DOCUMENT RESUME

ED 043 679 TM 000 157

AUTHOR	Rudegeair, Robert E.; Kamil, Michael L.
TITLE	Assessment of Phonological Discrimination in Children.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	TR-118
PUB DATE	Mar 70
CONTRACT	OEC-5-10-154
NOTE	20p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS	Audiclingual Skills, *Auditory Discrimination, Aural Stimuli, Discrimination Learning, *Grade1, *Kindergarten Children, Listening Comprehension, Listening Skills, *Listening Tests, *Phonemics, Verbal Communication
IDENTIFIERS	*Wepman Test of Auditory Discrimination

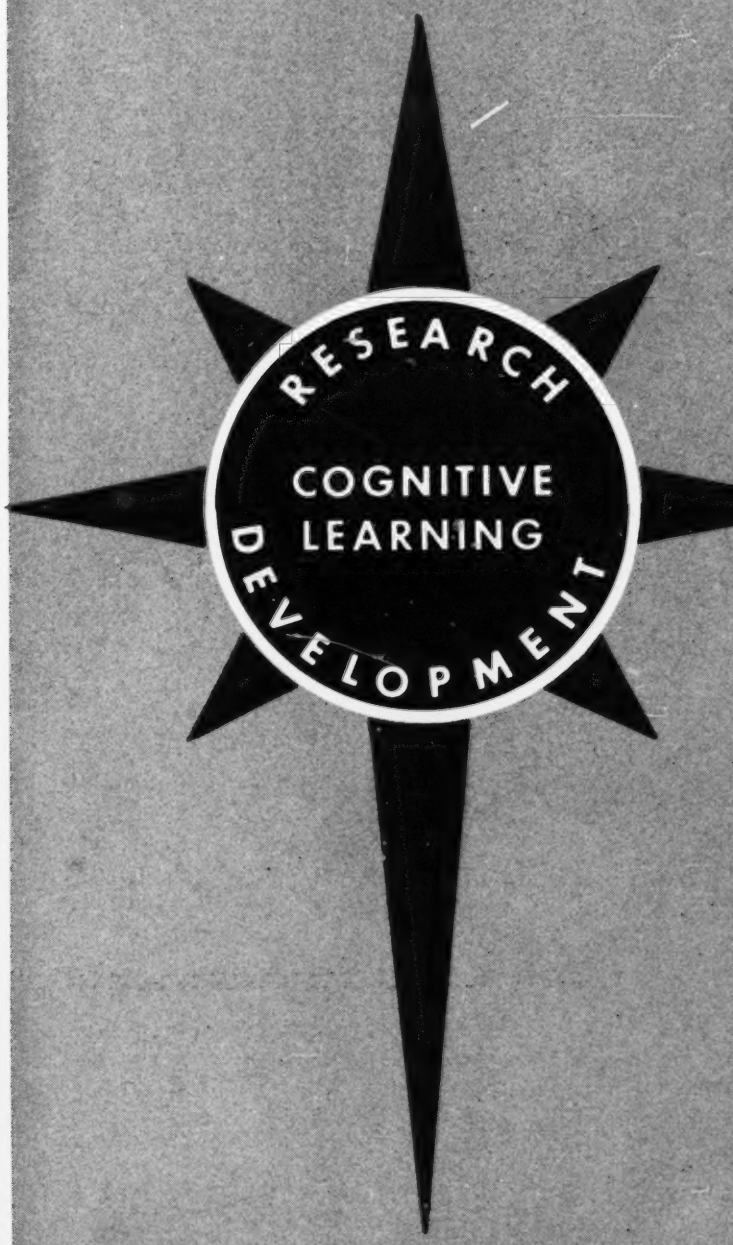
ABSTRACT

A review of the literature indicated that conventional tests are inadequate for accurate assessment of phonological discrimination ability in children. Higher error rates on discrimination tests than those which would be predicted from articulation measures seemed to implicate task variables. To reduce task difficulty, repeated contrast test pairs, consisting of CVC syllables in which the same phonemic contrast occurs twice, were developed. Multiple testing sessions were also employed to reduce initial task difficulty. Two experiments were conduced to determine the effects of task variables. The first experiment involved the use of nonsense syllables in an A-B-X paradigm. The subjects were 12 irst Grade and 12 Kindergarten children. The second experiment involved the use of real word items from the Wepman Test of Auditory Discrimination and the subjects were the First Grade group from experiment I and a new First Grade group of 12. Experiment I showed repeated constrast pairs easier to discriminate than initial or final contrast pairs and that results were poorer on the first day. In experiment II, first day performance was also significantly poorer than the other days. It is suggested that repeated testing is necessary for young children and that repeated contrast pairs may provide a means of obtaining more complete assessment of phonological discrimination ability in children. (Author/LR)

ASSESSMENT OF PHONOLOGICAL DISCRIMINATION IN CHILDREN

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR COGNITIVE LEARNING



+8

ED043679

Technical Report No. 118

ASSESSMENT OF PHONOLOGICAL DISCRIMINATION IN CHILDREN

By Robert E. Rudegeair and Michael L. Kamil

Report from the Project on Language Concepts and
Cognitive Skills Related to the Acquisition of Literacy

Robert C. Calfee and Richard L. Venezky
Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

March 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
Wage and Labor Standards Administration, U.S. Department of Labor, Washington

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research Division
Oregon State System of Higher Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

RESEARCH AND DEVELOPMENT CENTER POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

Russell J. Hosler
Professor, Curriculum and Instruction

Stephen C. Kleene
Dean, College of Letters and Science

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Clauston Jenkins
Assistant Director
Coordinating Committee for Higher Education

Donald J. McCarty
Dean
School of Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

Robert E. Grinder
Chairman
Department of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Ira Sharkansky
Associate Professor of Political Science

M. Crawford Young
Associate Dean
The Graduate School

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of Sociology

Robert E. Davidson
Assistant Professor, Educational Psychology

Russell J. Hosler
Professor of Curriculum and Instruction and of Business

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Anne E. Buchanan
Project Specialist
R & D Center

Frank H. Farley
Associate Professor, Educational Psychology

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

Robert G. Petzold
Associate Dean of the School of Education
Professor of Curriculum and Instruction and of Music

Robin S. Chapman
Research Associate
R & D Center

FACULTY OF PRINCIPAL INVESTIGATORS

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor of Educational Psychology

James Moser
Assistant Professor of Mathematics Education; Visiting Scholar

Richard L. Venezky
Assistant Professor of English and of Computer Sciences

Ted Czajkowski
Assistant Professor of Curriculum and Instruction

Lester S. Golub
Lecturer in Curriculum and Instruction and in English

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Alan Voelker
Assistant Professor of Curriculum and Instruction

Robert E. Davidson
Assistant Professor of Educational Psychology

John G. Harvey
Associate Professor of Mathematics and of Curriculum and Instruction

Milton O. Pella
Professor of Curriculum and Instruction (Science)

Larry Wilder
Assistant Professor of Curriculum and Instruction

Gary A. Davis
Associate Professor of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Thomas A. Romberg
Associate Director, R & D Center
Professor of Mathematics and of Curriculum and Instruction

Peter Wolff
Assistant Professor of Educational Psychology

M. Vere DeVault
Professor of Curriculum and Instruction (Mathematics)

Donald Lange
Assistant Professor of Curriculum and Instruction

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

MANAGEMENT COUNCIL

Herbert J. Klausmeier
Director, R & D Center
V.A.C. Henmon Professor of Educational Psychology

Thomas A. Romberg
Associate Director

James Walter
Director
Dissemination Program

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Development Program

* COMMITTEE CHAIRMAN

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Language Concepts and Cognitive Skills Related to the Acquisition of Literacy Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, this project's basic goal is to determine the processes by which children aged 4 to 7 learn to read, examining the development of related cognitive and language skills, and to identify the specific reasons why many children fail to learn to read. Later studies will be conducted to find experimental techniques and tests for optimizing the acquisition of skills needed for learning to read. By-products of this research program include methodological innovations in testing paradigms and measurement procedures; the present study is an example.

CONTENTS

		Page
	List of Tables and Figure	vii
	Abstract	ix
I	Introduction	1
	Background Literature	
	Conventional Methods of Assessing Phonemic Discrimination	1
	Attempts to Improve Conventional Testing	3
	The Present Study	4
II	Method	7
	Experiment I	7
	Materials	7
	Procedure	7
	Subjects	7
	Experiment II	7
	Materials	7
	Procedure	7
	Subjects	8
III	Results	9
	Experiment I	9
	Experiment II	10
IV	Discussion	13
	References	15
	Appendix	17

LIST OF TABLES AND FIGURE

Table		Page
1	Final Consonants Classes According to Their Influence on the Duration of the Preceding Vowel (from Peterson & Lehiste, 1960)	4
2	Complete List of Test Pairs Used for Consonant Discrimination Task	8
3	Percent Errors as a Function of Contrasts, Types of Items, and Grade	10
4	Error Rates for Individual Contrasts	11
5	Total Errors on ABA and ABB as a Function of Days and Grades	11

Figure		
1	Mean Errors per S as a Function of Days and Grade for all Test Items Combined	10

ABSTRACT

A review of the literature led the present investigators to conclude that conventional tests are inadequate for accurate assessment of phonological discrimination ability in children. Higher error rates on discrimination tests than those which would be predicted from articulation measures seemed to implicate task variables. To reduce task difficulty, the present investigators developed repeated contrast test pairs. Such pairs consist of CVC syllables in which the same phonemic contrast occurs twice (e.g., /bob/ - /dod/). Multiple-testing sessions were also employed to reduce initial task difficulty.

Two experiments were conducted to determine the effects of task variables. In Experiment I, nonsense syllables were used in an A-B-X paradigm. Subjects were 12 First Grade and 12 Kindergarten children. Mixed lists of repeated contrast, initial contrast, and final contrast pairs were tape recorded and presented to each S, one list per day for 6 days.

Experiment II involved using the real word items from the Wepman Test of Auditory Discrimination. Subjects were the 12 First Grade Ss from Experiment I and 12 additional First Grade Ss. Testing was conducted following the instructions for administration of the test. Four testing sessions were given each S.

The results for Experiment I showed that repeated contrast pairs were easier to discriminate than either initial or final contrast pairs and that there was no difference between initial and final contrast pairs. Performance on Day 1 was significantly poorer than on all other days, and there was no difference among the other 5 days.

For Experiment II, performance on Day 1 was also significantly poorer than on the other days. In addition, it was found that there was no difference in performance for "new" and "old" Ss.

The results have two major implications: First, repeated testing is a necessity for young children. Second, repeated contrast pairs may provide a means of obtaining a more complete assessment of phonological discrimination ability in children.

I
INTRODUCTION

BACKGROUND LITERATURE

Discrimination of speech sounds, in some sense, must precede the production of speech sounds. If a speaker can consistently articulate a given speech sound, he must be able to discriminate that sound from other sounds. Implicitly, investigators have assumed that phonological discrimination tasks measure the skills prerequisite to articulation. On the basis of this assumption, researchers have attempted to isolate discriminatory deficiencies and show that these deficiencies are related to articulatory problems. The purpose of this study is to test the validity of the instruments used in assessing speech sound discrimination ability and to develop ways of improving traditional tests.

The study of discrimination skills always involves inference. A task intervenes between behavior and observation. Consequently, it would seem that the task is a crucial factor in speech discrimination experimentation. A review of past research in which measurements of speech discrimination were undertaken is in order. The following survey concerns three common speech sound discrimination tests, the Travis and Rasmus Speech Sound Discrimination Test, the Templin Speech Sound Discrimination Test, and the Wepman Test of Auditory Discrimination. The bulk of research employing these tests has involved attempts to relate sound discrimination ability to articulation ability. Because such attempts are related to theoretical assumptions about discrimination ability, the research will be discussed in some detail.

Conventional Methods of Assessing Phonemic Discrimination

The Travis and Rasmus Speech Sound Discrimination Test (1931) represents an early attempt to assess discrimination ability and relate it to articulation ability. This test consists of 331 contrasting pairs (e.g., /ta/ - /da/) and 35 non-contrasting pairs (e.g., /ta/ - /ta/). Subjects, both adults and children in this case, were asked to make same or different judgments in response to the pairs as they were presented orally. Three hundred pairs involve consonants, and 66 pairs involve single vowels. All consonant contrasts occur in consonant-vowel (CV) nonsense syllables, the vowel being /a/. The practical advantage of a test of this size is questionable. Since the Travis and Rasmus Test, researchers have only used "minimally" different speech sounds in contrast pairs.

Travis and Rasmus (1931) compared Ss with normal articulation and Ss who had mild-to-severe functional disorders of articulation. Four age levels were studied: adults, Fourth and Fifth Graders, Second and Third Graders, and junior primary and First Grade pupils. The investigators concluded that Ss with normal articulation are superior in speech sound discrimination skill to articulatory defectives at all age levels. The important point, for our purposes, is not the conclusion itself, but the data on which the conclusion was based. Control Ss with normal articulation scores in the junior primary and First Grade group (n = 50) made an average of 8.2 errors on the discrimination test. Since less than 10% of the contrast pairs included on the Travis and Rasmus Test are what are known as "minimal" contrasts (i.e., contrasts involving only single-feature changes), this error rate seems surprisingly high in comparison to what is known about articulation abilities for this age level (Venezky & Calfee, 1968). Throughout the ensuing survey, an attempt will be made to emphasize the size of the error rates obtained in speech sound discrimination measures for young children.

The Travis and Rasmus Speech Sound Discrimination Test was one of a battery of tests

used by Hall (1938). Her purpose, like that of Travis and Rasmus, was to compare speech defectives with normal speakers. Subjects were university freshmen and elementary school children. In direct opposition to the findings of Travis and Rasmus, Hall found that functional articulatory speech defectives, at both age levels, showed no inferiority to normal speakers on any of her measures. The elementary school control Ss (N = 64; age = 7 through 13 years) in Hall's study made a mean of 15.10 errors on the discrimination test. The fact that this is nearly twice as large as the error rates found by Travis and Rasmus for junior primary and First Grade Control Ss is strange. Just as strange is Hall's finding of an average of 8.4 errors for the university freshmen control Ss. This mean is almost identical to that of the junior primary and First Grade control Ss in the Travis and Rasmus study. Hall explains that this discrepancy may have arisen because of experimental group selection procedure, experimenter variation, or, most importantly, because Travis and Rasmus tested control Ss in groups and experimental Ss in groups and experimental Ss individually, while both experimentals and controls in the Hall study were tested in the same groups.

The Travis and Rasmus Test was replaced when it became apparent that only minimally different speech sounds should be incorporated in discrimination tasks. Templin (1943), in an an attempt to produce a useable, "short test" of sound discrimination, devised a 70-item test consisting of 51 contrasting pairs and 19 non-contrasting pairs. Although Templin's Test used nonsense syllables and relies on same-different judgments, it is different, in several respects, from the Travis and Rasmus model. While Travis and Rasmus used only CV syllables to test consonant discrimination, Templin used CV, VC, and VCV syllables. She also utilized three vowels and a diphthong in constructing the syllables (/i/, /e/, /o/, and /ai/). Furthermore, no vowel contrasts are included in the Templin Test.

Subjects for the Templin Test were 301 pupils attending Second through Sixth Grade. Subjects showed a mean of 18.29 errors, or an overall error rate of 26%. The error rate may represent only a lower limit, since Templin seems to have lumped together both "same" and "different" errors in the figures she reported.[1] In any case, the error rate is well above what the present authors feel would be an accurate picture of sound discrimination abilities of this age group. The short test discussed here and a shorter version consisting of 50 same and different items (Templin, 1957) have been utilized by several investigators who were concerned with the relation between articulation abilities and phonemic discrimination abilities.

Kronvall and Diehl (1954) matched a group of 6- to 9-year-old speech defectives (N = 30) with a normal group (N = 30). Using the 70-item Templin Speech Sound Discrimination Test, they found that the speech defective group made significantly more discrimination errors. Cohen and Diehl (1963) replicated the Kronvall and Diehl study and confirmed the results. These two studies yielded 12.37 and 12.13 mean errors, respectively for control Ss. Both figures constitute a 17% overall error rate.

Sherman and Geith (1967) reversed the normal procedure in selecting Ss for their study. They first administered the Templin Speech Sound Discrimination Test to 529 Kindergarten pupils; then they selected 18 high-scoring Ss and 18 low-scoring Ss in order to compare the two groups on an articulation test (Templin-Darley Picture Articulation Test). The investigators used this procedure because they felt that other studies, which did not find significant differences between defectives and normals, failed to use appropriate selection procedures. Subjects in previous studies were chosen on the basis of their articulation deficiencies and consequently were heterogeneous with respect to the etiology of the speech disorder. Results of the Sherman and Geith study showed a significant relationship between high discrimination errors and high articulation errors. The investigators concluded that deficient discrimination skill is causally related to poor articulation. No overall error rates or item analyses from the sound discrimination test are available for this study.

The Wepman Auditory Discrimination Test (1958) has also been relied upon as an instrument in correlational studies. Wepman's Test contains 40 real-word pairs—10 "same" pairs and 30 "different" pairs. Only errors on

---

[1] It has been shown that performance on "same" pairs is significantly better than performance on "different" pairs (Briere, 1967). If "same" pairs are included in calculating error rates, the estimate will be lower than estimates based solely on "different" pairs. Thus, if Templin counted all errors, a corrected estimate (based only on "different" pairs) would be greater than the 26% cited.

"different" pairs are scored. "Same" pairs were included to serve as a check on the validity of the test (Wepman, 1960). Both members of a contrasting pair are equated for frequency of occurrence according to the Thorndike-Lorge count (1944). All consonant contrasts on the test constitute changes in place of articulation only. In addition to consonant contrasts, four vowel contrasts appear on each of the two test forms available. Wepman (1960), in a preliminary report on the findings of studies using his test, concluded that there was a definite relation between poor articulation and poor discrimination but no data were given.

Using the Wepman Test, Prins (1963) approached the traditional question in a new and interesting way. He argued that attempts to correlate a gross articulation score with an overall speech sound discrimination score were unfruitful; rather, if a relation between the two skills exists, then a relation between specific deviations of speech sound production and phonemic discrimination should be demonstrable.

Prins obtained articulation measures on First Grade pupils, and selected experimental Ss on the basis of more than three errors. Each S's errors were categorized by type. Then the Wepman Test of Auditory Discrimination was administered. It was found that children in the experimental group whose articulation errors were phonemic substitutions involving the place feature tended to make many errors on the Wepman Test. (Recall that Wepman Test items all involve place contrast.) Those children in the experimental group whose substitution errors differed greatly from the target phoneme made fewer errors on the Wepman Test. There was no significant difference between the Wepman scores of the experimental group and those of a control group chosen on the basis of few articulatory errors. The control Ss showed a mean of 5.58 errors on the Wepman Auditory Discrimination Test; this constitutes an overall error rate of 18.6%.

Attempts to Improve Conventional Testing

No speech sound discrimination error rate for normal groups reported in the literature was less than the 17% reported by Kronvall and Diehl (1954). Since the assumption was made that speech sound discrimination tests should be measuring a skill prerequisite to articulation, the error rates cited are unusually high. It would seem reasonable to conclude that conventional tests are not measuring discrimination skill alone. Thus, the validity of these tests is questionable. A research program was initiated at the Wisconsin Research and Development Center to isolate variables affecting the measurement of sound discrimination. Four experiments are reported as representative of the experiments conducted. Details of methodology have been omitted from the following account. The Appendix is a summary of the procedures employed in each of the experiments. In general, each experiment constituted an attempt to produce a speech sound discrimination test that would allow valid comparisons of discrimination and articulation errors.

For the first experiment, the instrument used was designed to avoid the overt same-different procedure employed by Templin (1943) and Wepman (1958). It was felt that some Ss in this age group (5-6 years) were unable to comprehend the relevant dimensions required in making same-different judgments. Furthermore, the A-X paradigm (i.e., same-different procedure) has recently been criticized by Briere (1967). He shows evidence that "an A-X paradigm in language perception might simply be a measure of the S's bias to respond 'same' or 'different'." Also, error rates for "same" pairs are significantly lower than error rates for "different" pairs.

CVC nonsense syllables were employed in an A-B-X paradigm[2] in an attempt to control for differential familiarity of test items. All contrasts employed in this test were initial consonant contrasts. Subjects showed an average of 5.7 errors. This constitutes an overall error rate of 18.9%.

The second experiment was, in purpose and design, similar to the first. Final consonant contrasts in CVC nonsense syllables were employed. Subjects made an average of 7.3 errors, or an overall error rate of 24.5%. Obviously, the goal of deflating error rates was not realized in the first two experiments. In the third experiment, attention was turned to the relationship between articulation and discrimination.

An imitative articulation test similar to Templin's (1947) was devised and administered to Kindergarten pupils. An experimental

---

[2] An A-B-X paradigm consists of presenting word A, then word B, and finally the question "who said X?", where X is either A or B. In an A-X paradigm, word A is presented, followed by word X, where X is either A or another word (B). This is the traditional same-different paradigm.

group was formed from Ss making more than 14 articulatory errors. From the remaining Ss, an equal number of control Ss was randomly selected. The two groups were then given a phonemic discrimination test consisting of initial and final contrasts. Both groups showed a mean of 6.3 errors, or an overall error rate of 31.6%. Since the error rates for both groups were identical, no further analyses were conducted. It was felt that such a high error rate on what is presumed to be a simple task left only the task itself to be blamed. A retesting of a randomly selected portion of the experimental Ss one week later reinforced this idea. Subjects were inconsistent in the type of errors made.

A final experiment in this series, reported by Skeel, Calfee, and Venezky (1969), tested fricative discrimination in preschool children. Results showed, as expected, that Ss in this age group (3 years 11 months to 5 years 1 month) had considerable difficulty in discriminating fricatives from one another. A 28% overall error rate was obtained. Subjects were also tested on articulation of the same fricative set. The correlation between poor articulation ability and poor discrimination was significant ($R = .62$, $p < .01$). Item analysis, however, showed that articulation error patterns differed from discrimination error patterns. This discrepancy points out the danger of using gross scores in correlating articulation and speech discrimination ability. If speech sound discrimination errors are not predictable from a S's articulation errors, then the correlation if obtained, is not meaningful. It is possible that both articulation ability and phonemic discrimination ability are related to some third variable (e.g., general language processing ability).

The above studies represent a sample of attempts to manipulate variables in speech discrimination tasks in order to reduce error rates. The attempts have failed to show any improvement over previous experimentation in this area. The variables that were manipulated in the experiments so far are not crucial in affecting task difficulty. If the assumption that task variables are interfering with accurate measurement of sound discrimination in children is true, then the identification of these variables is vital.

THE PRESENT STUDY

Assuming that, for young children, there is an inherent difficulty in normal phonemic discrimination tasks, major modifications in design and administration seem necessary. Phonemic discrimination tasks usually involve minimally different pairs of syllables (e.g., /mom/ - /nom/). One way of making the task less difficult would be to increase the discriminability of items in a pair. It is desirable, though, to control closely the contrasts so that specific inferences can be made about the cues Ss are using in discrimination.

To solve the problem, we developed test items called repeated contrast pairs. A repeated contrast pair consists of two CVC syllables in which the same phonemic contrast occurs twice. For example, /p æ p/-/t æ t/ is a repeated contrast pair in which the phonemic contrast /p/-/t/ occurs in both initial and final positions. Subjects must use the phonemic information to discriminate correctly but they are given the same information twice.

Vowel length has been found to vary as a function of the following consonant (Peterson & Lehiste, 1960). Skeel, Calfee, and Venezky (1969) suggest that Ss might be using these varying durations as cues to the discrimination of contrasting elements in final position. They found that errors were higher on CV syllables than on VC syllables involving the same contrasts. Thus, repeated contrast pairs might include vowel duration cues in addition to other phonemic cues. Peterson and Lehiste (1960) identified four categories of consonants within each of which vowel length is not differentially affected. The categories are given in Table 1. For CVC syllables, the consonants within a given category have the same effect on the duration of the vowel. Choosing consonant contrasts from within these categories will eliminate potential vowel duration cues. Discrimination, then, can only be made on the basis of the contrasting consonant elements.

Table 1

Final Consonants Classed According to Their Influence on the Duration of the Preceding Vowel (from Peterson & Lehiste, 1960).

Range of Representative Vowel Duration in Centiseconds	Consonants
19-21	/p, t, k, č/
26-28	/f, θ, s, š/
30-32	/b, d, g, ǰ, m, n/
37-41	/v, ð, z, ž/

In a recent criticism of the use of an A-B-X paradigm in phonological testing (Briere, 1967), it was found that the A-B-A alternative produces significantly more errors than the A-B-B alternative. In light of Briere's additional criticism of the A-X paradigm, the choice of a suitable task involves a dilemma. One solution is to use the A-B-X paradigm, but carefully counterbalance presentations of all items so that contrasts occur equally in A-B-A and A-B-B instances. This procedure was adopted for the present study.

Of the studies reported above, only Skeel, Calfee, and Venezky (1969) involved training Ss in the discrimination task. In that study, no significant effect of practice was found but the 1-week intervals between sessions may have mitigated the effectiveness of the training. For the present study, these were important factors to be considered.

The specific hypotheses to be tested in the present study are: 1) that the use of repeated contrast items will yield lower error rates than conventional pairs, and 2) that multiple testing will produce more stable performance in a discrimination task of this sort. Both hypotheses represent attempts to reduce task difficulty without sacrificing control over test procedures.

II
METHOD

EXPERIMENT I

Materials

Using the categories given in Table 1, 21 repeated contrast pairs (e.g., /bob/ - /dod/) were constructed. For each phonemic contrast used in a repeated contrast pair, a minimal initial pair (e.g., /bob/ –/dob/) and a minimal final contrast pair (e.g., /bob/ - /bod/) were also constructed. Of the 21 contrasts, 13 were composed of consonants which were drawn from the same vowel length category (see Table 1). The remaining eight contrasts were voiced-voiceless pairs (e.g., /p æ p/ - /b æ b/) drawn from between categories. A list of all test pairs is given in Table 2.

From the pairs in Table 2, six lists were formed. Each pair appeared once in the first three lists and once in the second three. Each list was composed of 24 test pairs, 7 each of repeated, initial, and final contrasts in randomized orders and 3 real word pairs (e.g., car-dog) at the end of the list. The real word pairs served as control pairs. Voicing pairs were interspersed with controlled vowel length pairs in the lists.

All items were recorded on an Ampex stereo tape recorder at the WHA Radio Studios on the UW campus. The speaker's dialect is best described as Upper Midwestern.

Procedure

An A-B-X paradigm was used. A ready signal (bell) sounded on the left speaker, followed by item A of a test pair on the same speaker. Item B of the pair was presented on the right speaker. The question "Who said X?" occurred on both speakers. The S's task was to match X to A or B by pointing to the appropriate speaker. Informative feedback was then presented over the correct speaker: "I said X." The next trial was signaled by the bell.

Tapes were played on an Ampex stereo tape recorder, Model 1100. The speakers were placed 6 feet apart and directly to the right and left of the Ss.

Subjects were tested six times, once a day for about 5 minutes per day. Lists were arranged in counterbalanced sequences across Ss.

Subjects

Twelve Kindergarten and 12 First Grade Ss from a Madison public school participated. The mean ages for the two groups were 5.5 and 6.5 years, respectively.

Responses were recorded by E on prepared data sheets.

EXPERIMENT II

Materials

Test pairs used in this experiment were those from the Wepman Test of Auditory Discrimination (WTAD) (1958). Two lists were used, Form I and Form II of the WTAD.

Each form contains 30 minimally contrasting real word pairs (i.e., "different" pairs) and 10 non-contrasting real word pairs (i.e., "same" pairs). There are 13 consonant contrasts involving change in place of articulation alone and 4 vowel contrasts. Each of the 13 consonant contrasts appears once in initial position and once in final position on each form.

Each form consists of 26 consonant contrasts, 4 vowel contrasts, and 10 non-contrasting pairs, for a total of 40 pairs.

Procedure

Presentation of the items conformed to the instructions given for administration of the

Table 2

Complete List of Test Pairs Used for Consonant Discrimination Task

Repeated Contrast Pairs	Minimal-pair (initial)	Minimal-pair (final)
pæp - tæt	pæp - tæp	pæp - pæt
tæt - kæk	tæt - kæt	tæt - tæk
tæt - čæč	tæč - čæč	čæt - čæč
fæf - θæθ	fæf - θæf	fæf - fæθ
θæθ - sæs	θæs - sæs	sæθ - sæs
sæs - šæš	sæs - šæs	sæs - sæš
bob - dod	bob - dob	bob - bod
dod - gog	dog - gog	god - gog
dod - ǰoǰ	dod - ǰod	dod - doǰ
mom - non	mom - nom	mom - mon
vov - ðoð	vov - ðov	vov - voð
ðoð - zoz	ðoð - zoð	ðoð - ðoz
zoz - žož	zoz - žoz	zoz - zož
pæp - bæb	pæp - bæp	pæp - pæb
tot - dod	tod - dod	dot - dod
kæk - gæg	kæg - gæg	gæk - gæg
fæf - væv	fæf - væf	fæf - fæv
θæθ - ðæð	θæθ - ðæθ	θæθ - θæð
sos - zoz	sos - zos	sos - soz
šoš - žož	šoš - žoš	šoš - šož
čoč - ǰoǰ	čoč - ǰoč	čoč - čoǰ

WTAD. From a position behind Ss, E read each pair aloud. Subjects were required to respond "yes" if the items in a given pair were the same and "no" if they were different. Errors were recorded on prepared data sheets. The E was the same speaker who prepared the tapes for use in Experiment I.

Subjects

Two groups of 12 First Grade Ss took part. The mean age for Group I was 6.8 years; for Group II the mean age was 6.9 years. Mean percentile scores on the Metropolitan Readiness Test were 76.8 for Group I and 81.5 for Group II.

Each S in the first group was tested four times, once a day in individual sessions. Presentation of Lists I and II was counterbalanced over Ss with one restriction: On Day 1, a particular S received either Form I or II. That S then received the remaining Form on Day 2. The procedure was repeated on Days 3 and 4.

Group II consisted of the same 12 Ss who participated in Experiment I. They were tested only once, with Ss assigned randomly, six to Form I, six to Form II.

III
RESULTS

EXPERIMENT I

Overall, the error rate for Experiment I was 13.9%. For the Kindergarten Ss, the error rate was 17.7%, while it was 10.2% for the First Grade Ss. When divided by item types, final contrast pairs had an error rate of 16.6%, initial contrast pairs had an error rate of 15%, and repeated contrast pairs showed an error rate of 9.9%.

Errors were tabulated for each S for all 6 days. Two preliminary analyses were performed. First, the errors on the 13 controlled vowel-length pairs were analyzed in a repeated measures analysis of variance, Ss nested within grades and crossed with contrast types (repeated, final, or initial) and days. Corrections for repeated measures were made, where appropriate, by the Geisser and Greenhouse method (1958). There was a significant effect of contrast types, $F(1, 22) = 6.83$, $p < .025$. Subsequent tests by the Newman-Keuls method showed a significantly lower mean number of errors on repeated contrast pairs than on either type of single contrast pair ($p \leq .05$). There was no difference in mean errors between minimal final and initial pairs ($p > .05$). The data for the eight voicing contrasts were analyzed in the same manner. Again, the effect of items was significant, $F(1, 22) = 5.29$, $p < .05$. Subsequent tests again showed a lower mean number of errors for repeated contrast pairs than for either type of single contrast pair ($p < .05$). Literally, no difference between final and initial contrast pairs existed, as the means were identical. In both analyses, no interactions were significant, so for the final analysis the errors for all items combined were analyzed by a similar repeated measures design. All main effects were significant: days ($F(1, 22) = 11.32$, $p < .01$); contrast types ($F(1, 22) = 13.21$, $p < .01$); and grades ($F(1, 22) = 5.23$, $p < .05$). None of the interactions reached significance.

Subsequent tests by the Newman-Keuls method showed that the mean number of errors per S on repeated contrast pairs (4.2) was significantly smaller than either the mean errors per S on initial pairs (6.4), or the mean errors on final pairs (7.0), $p < .01$. There was no difference between the means for initial and final pairs ($p > .05$).

Mean errors per S as a function of days is plotted in Fig. 1. Because there is no appropriate statistic for evaluating the differences between levels of an ordered variable, the Newman-Keuls test was used. Only the differences between Day 1 and all other days were significant ($p < .01$).

Because Day 1 performance is significantly worse than the performance on Days 2-6, error rates were calculated for the last 5 days. Overall, the error rate is 12.1%. For Kindergarten, the error rate is 15.7%; for First Graders, it is 8.5%. For final, initial, and repeated contrast pairs, the error rates are 14.3%, 14.0%, and 8.0%, respectively.

Table 3 shows error rates separated by voicing contrasts versus others. In all cases, voicing contrasts show a slightly lower error rate.

Separate item analyses were also performed. Error rates separated by item types for all contrasts are given in Table 4. Three contrasts, /f/ - /θ/, /v/ - /ð/, and /m/ - /n/, account for 43% of all errors on controlled vowel-length pairs.

One final analysis was conducted on the data from Experiment I. The A-B-A and A-B-B errors for each S for each day were analyzed. Table 5 gives the number of A-B-A and A-B-B errors as a function of days and grades. A repeated measures analysis of variance showed no significant difference between the mean errors on A-B-A and A-B-B ($F(1, 22) = 3.35$, $p > .05$). There was a significant difference between the means for the interaction of A-B-A and A-B-B errors with days ($F(1, 22) = 5.43$,

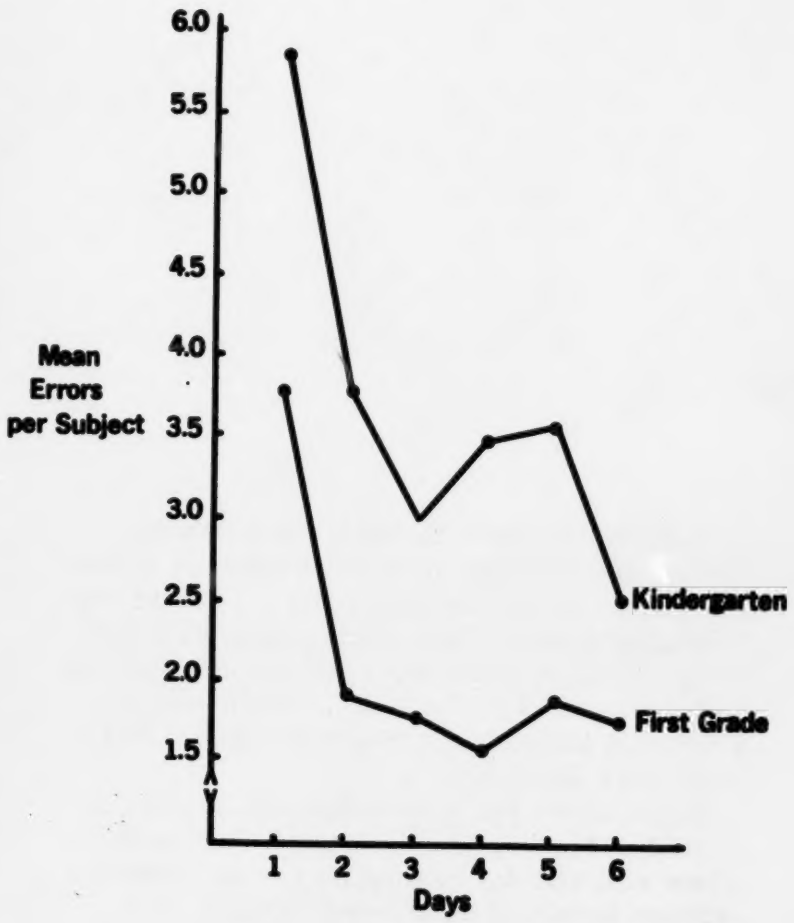


Figure 1. Mean Errors per S as a Function of Days and Grade for All Test Items Combined

$p < .05$). Subsequent t tests between the A-B-A and A-B-B error means on each of the six days revealed significantly more A-B-A errors only on Day 1, $t(22) = 4.706$, $p < .001$. There were no significant differences on subsequent days.

EXPERIMENT II

For this experiment, the overall error rate was 11%, excluding errors made on "same" pairs. Only 22 errors were made on "same" pairs, for a 5% error rate.

Errors were tabulated for each S in Group I for each day. Again, "same" errors were excluded. A subjects x days repeated measures analysis of variance was performed on the data. There was significant improvement over days, $F(1, 11) = 7.10$, $p < .05$, after correction for repeated measures. The Newman-Keuls method was again used to evaluate differences between the means for the 4 days. Again, only the difference between Day 1 and all other days was significant ($p < .05$).

As in Experiment I, the error rate was calculated after eliminating Day 1 errors. The reduced error rate is 8.7% (again, excluding "same" errors).

Since Group II was only tested once, the errors for each S were tabulated for Day 1. A t test was performed on the Day 1 errors from Group I and the errors from Group II. Although the mean number of errors for Group I on Day 1 was 5.4 and the mean number of errors for Group II was 2.9, the difference was not significant, $t(22) = 1.89$, $p > .05$.

Table 3

Percent Errors as a Function of Contrasts, Types of Items, and Grade

Voicing Contrasts				Controlled-Vowel Contrasts			
	K	G1	All Ss		K	G1	All Ss
Redundant	12.5	4.7	8.6	Redundant	15.1	6.4	10.2
Minimal Initial	19.3	10.4	14.9	Minimal Initial	19.6	11.5	15.5
Minimal Final	18.7	10.9	14.8	Minimal Final	20.2	15.4	17.8

Table 4

Error Rates for Individual Contrasts

Controlled Vowel Length Contrasts		Voicing Contrasts	
Contrast	Error Rate	Contrast	Error Rate
p - t	24%	p - b	15%
t - k	12%	t - d	19%
t - č	24%	b - g	18%
f - θ	74%	f - v	26%
θ - s	37%	θ - ð	37%
s - š	18%	s - z	32%
b - d	17%	š - ž	32%
d - g	11%	č - ǰ	24%
d - ǰ	17%		
m - n	47%		
v - ð	43%		
ð - z	25%		
z - ž	33%		

Table 5

Total Errors on ABA and ABB as a Function of Days and Grades

	Kindergarten		Grade 1	
	ABA	ABB	ABA	ABB
1	45	25	31	16
2	29	17	12	11
3	20	16	11	11
4	24	18	6	13
5	21	22	12	11
6	17	13	12	9
Total	156	111	84	71

GPO 817-447-4

IV
DISCUSSION

For Experiment I, the major results indicate: (1) performance improves significantly after the first day of testing; (2) repeated contrast pairs are easier to discriminate than either initial contrast pairs or final contrast pairs; (3) a small set of contrasts account for a large proportion of the errors.

It was argued earlier that task variables have inflated discrimination measures in previous studies. The improvement over days is strong support for this position. Obviously, Ss do not learn to discriminate phonemes between the first two sessions. The significant improvement may be attributed to an increased facility with the task, or to factors of attention, or to both.

One specific example of such improvement is the reduction of errors in response to A-B-A alternatives. On Day 1, there are significantly more A-B-A errors than A-B-B errors. This is in complete agreement with the findings of Briere (1967). However, there are no significant differences between A-B-A and A-B-B errors on any of the succeeding days. What Briere attributed to limitations in auditory memory span seems to be merely the result of attentional processes or some simple form of task learning.

While the reduction of A-B-A errors from Day 1 to Day 2 is large, it does not account for all of the improvement in performance. Thus, it may be speculated that there is some more global mechanism generating improvement to which the reduction of errors on A-B-A pairs is related.

Most earlier studies assumed that Ss were attending to the critical attributes of stimuli as a matter of course. In the present study, superior performance on repeated contrast pairs implies that Ss may not always receive the critical information to discriminate conventional minimal pairs. Repeated contrast pairs provide Ss more than one opportunity to obtain the necessary information, thus demanding less of the child in terms of attention. Consequently, a truer picture of discrimination ability in children can be obtained, uncontaminated by attentional factors.

The finding that three contrasts (/f/ - /θ/, /v/ - /ð/, /m/ - /n/) accounted for 43% of all errors in Experiment I is not unusual or surprising. Such a finding has been reported by other investigators (Travis and Rasmus, 1931; Templin, 1943; Skeel, Calfee, & Venezky, 1969). It is predictable from a knowledge of the impoverished acoustic cues involved. The fricative contrasts, /f/ - /θ/ and /v/ - /ð/ are distinguished, not by their acoustic energy spectra, but only by formant transitions, in contrast to other fricatives (Delattre, Liberman, & Cooper, 1962). Further, the six phonemes in question are marked by low intensity. It may also be that the auditory presentation precludes using important visual cues normally present with these phonemes, since they all involve labial or dental movements.

Another finding of interest is that, overall, there was no difference in performance on final and initial contrast pairs. This is in contrast to earlier studies reporting that final pairs were significantly more difficult to discriminate (Templin, 1957).

Where voicing contrasts are concerned, vowel duration as a cue is confounded with voicing as a cue for repeated and final pairs. Initial consonants, though, have no effect on the length of vowels in CVC's (Peterson & Lehiste, 1960). Thus, the initial contrast pairs for voicing contrasts are, in effect, controlled vowel-length pairs, since both CVC's in a pair end with the same consonant. Initial voicing contrasts serve as an adequate control in determining the effect of vowel duration as a cue. There was no difference between initial and final contrast pairs in the analysis of the controlled vowel-length minimal pairs. To conclude that Ss use vowel

duration as an additional cue in discrimination, final contrast pairs must exhibit fewer errors than initial pairs in voicing contrasts. Since no significant difference was found, there is no indication that Ss make effective use of vowel duration in discriminating CVC's differing in the final consonant.

The generality of the statement that performance improves after Day 1 is supported by the results of Experiment II using the Wepman Test of Auditory Discrimination. Although the task, materials, and presentation differ from those used in Experiment I, the finding with respect to repeated testing is the same. In addition, this finding also supports the notion that general task variables can and do inflate error rate in conventional testing of speech sound discrimination.

The use of Group II, the same First Grade Ss previously tested in Experiment I, was an attempt to measure the amount of transfer from previous discrimination testing. Although the number of errors for Group II was 35 as opposed to 65 for Group I on Day 1, the difference was not statistically significant. Thus, while there does seem to be some evidence of carry-over, it is perhaps overly optimistic to expect that it should be reliable after a 5-month interval.

The experiments reported above indicate the need for several important modifications in methodology for optimal assessment of phonemic discrimination in young children. Repeated testing is a necessity. Since stable performance is reached after Day 1, purer measures can only be obtained in subsequent sessions. No study in this area has ever discarded results from an initial testing session. Therefore, conclusions in previous studies are based on error rates inflated by confounding factors and only partially related to discrimination ability.

The use of repeated contrast pairs in the present study represents an additional step toward attaining purer discrimination measures. Researchers have relied on normal minimal pairs without questioning their validity in testing young children. Since the same control over contrasting elements is available with repeated contrast pairs, their use will sharply increase the accuracy of item or S-related measures.

Several other aspects of the testing format used in Experiment I may prove to be desirable modifications of speech discrimination tasks. First, taped presentations permitted control of items across Ss, a precaution particularly necessary in diagnostic use of tests. Second, the choise of CVC nonsense syllables avoided the potential problems of familiarity inherent in the use of real words. Finally, the A-B-X paradigm required minimal instruction in comparison to same-different tests; Ss' responses were not dependent on concepts of "same sounds" or "different sounds."

The value of future research in the area of phonemic discrimination in children seems dependent on improvement of the testing instrument; repeated testing and the use of repeated contrasts to control for fluctuating attention would appear to be important steps in this direction. Certainly, past research must be re-evaluated in light of the present experimental findings.

REFERENCES

Briere, E. J. Phonological testing reconsidered. Language Learning, 1967, XVII, 163-171.

Cohen, J. & Diehl, C. Relation of speech-sound discrimination ability to articulation-type defects. Journal of Speech and Hearing Research, 1963, 28, 556-560.

DeLattre, P. C., Liberman, A. M., & Cooper, F. S. Formant transitions and Loci as acoustic correlates of place of articulation in American fricatives. Studia Linguistica, 1962, 16, 104-121.

Geisser, S. & Greenhouse, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 1958, 29, 885-891.

Hall, M. Auditory factors in functional articulatory speech defects. Journal of Experimental Education, 1938, 7, 110-132.

Kronvall, E. L. & Diehl, C. F. The relation of auditory discrimination to articulatory defects of children with no known organic impairment. Journal of Speech and Hearing Disorders, 1954, 19, 335-338.

Peterson, G. E. & Lehiste, I. Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 1960, 32, 693-703.

Prins, D. Relations among specific articulatory deviations and responses to a clinical measure of sound discrimination ability. Journal of Speech and Hearing Disorders, 1963, 28, 382-388.

Sherman, D. & Geith, A. Speech sound discrimination and articulation skill. Journal of Speech and Hearing Research, 1967, 10, 277-280.

Skeel, M. Calfee, R. C., & Venezky, R. L. Perceptual confusions among fricatives in preschool children. Technical Report from the Research and Development Center for Cognitive Learning, The University of Wisconsin, 1969, No. 73.

Templin, M. A study of sound discrimination ability of elementary school pupils. Journal of Speech Disorders, 1943, 8, 127-132.

_____. Certain Language Skills in Children: Their Development and Interrelationships. Minneapolis: University of Minnesota Press, 1957.

Thorndike, E. L. & Lorge, I. The Teachers Word Book of 30,000 Words. New York: Columbia University Press, 1944.

Travis, L. E. & Rasmus, B. The speech sound discrimination ability of cases with functional disorders of articulation. Quarterly Journal of Speech, 1931, 17, 217-226.

Venezky, R. L. & Calfee, R. C. The patterning of articulation errors in kindergarten and first grade. Paper presented at Psycholinguistics and Reading Presession, International Reading Association, Boston, April, 1968.

Wepman, J. M. Auditory Discrimination Test (Manual of Directions). Chicago: Language Research Associates, 1958.

Wepman, J. M. Auditory discrimination, speech and reading. Elementary School Journal, 1960, 60, 325-333.

GPO 817-447-3

APPENDIX

SUMMARY OF PROCEDURES USED IN ATTEMPTS TO IMPROVE CONVENTIONAL TESTING

In the first experiment, 25 Kindergarten pupils attending a public school in Madison, Wisconsin, served as Ss, whose ages ranged between 5 and 6 years. Thirty contrast pairs were constructed. All contrasts were initial consonant contrasts in CVC syllables (e.g., /ʃif/ - /sif/). Place of articulation, manner of articulation, and voicing were manipulated. Stops, fricatives, and affricates were included. Each S received all 30 contrasts. An MTA scholar teaching machine was modified for presentation of the test items. A description of the technique employed can be found in Skeel, Calfee, and Venezky (1969). The child faced a visual display while the test items were presented in an A-B-X paradigm. S responded by pushing a panel that represented the source of the correct utterance.

In the second experiment, 57 Kindergarten pupils between 5 and 6 years of age served as Ss. All attended a public school in Madison, Wisconsin. The only controlled difference between this experiment and the first is the position of the contrasting elements in the CVC syllable. In this study, only final contrasts were tested.

For the third experiment, 97 Kindergarten pupils attending public schools in Madison, Wisconsin, served as Ss. Their ages ranged between 5 and 6 years. A list of 20 minimally distinct CVC nonsense syllables was constructed. Both initial and final contrasts were included in the test. Stops, fricatives, affricates, and nasals (/m/ and /n/) comprised the contrasting elements. The technique using the modified teaching machine was abandoned in this experiment. An Ampex 1100 stereo tape recorder was used in presenting the stimulus material. Left and right speakers were placed about six feet apart in front of S, who sat mid-way between them. A warning signal was heard on the left speaker, immediately followed by the first member of the contrast pair. Approximately one second later, the second member of the pair was presented over the right speaker. One second later "who said X" was presented over both speakers. Ss responded simply by pointing to the appropriate speaker. After a three second interval which allowed time for the response, "I said X" was heard over the appropriate speaker. A warning signal initiated the next cycle.

For the final experiment, 17 Ss, ranging in age from 3 years 11 months to 5 years 1 month, were presented a total of 90 CV and VC contrast pairs over a 7-week period. Contrast pairs consisted of six fricatives (/f/, /v/, /θ/, /ð/, /s/, and /z/), each contrasted with each of the other five, initially and finally, and in combination with three syllable nuclei (/a/, /i/, and /aɪ/). In the first week, Ss were given a pretraining task. For each of the subsequent six weeks, 15 contrast pairs were presented along with seven control pairs (interspersed in the list). These control items remained constant for all sessions. The modified MTA technique, referred to earlier in the description of the first experiment, was used.

GPO 817-447-2